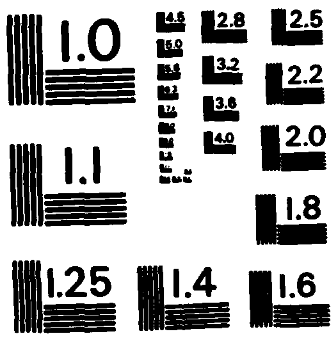
AD-A159 926 ESTIMATION OF POPULATION PROPORTION AND OF PROBABILITIES OF MISCLASSIFICA..(U) MARYLAND UNIV COLLEGE PARK DEPT OF MANAGEMENT SCIENCES AND ST.. 1/1

UNCLASSIFIED N L JOHNSON ET AL. AUG 85 UM/MSS/85/7 F/G 12/1 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(2)

ESTIMATION OF POPULATION PROPORTION AND OF PROBABILITIES OF MISCLASSIFICATION FROM IMPERFECT BINOMIAL DATA BY COMPARISON WITH RESULTS OF A STANDARD METHOD OF KNOWN INPERFECTION

Norman L. Johnson
University of North Carolina
Chapel Hill, North Carolina

and

Samuel Kotz
University of Maryland
College Park, Maryland

N00014-84-K-0301

August 1985

-A-

ESTIMATION OF POPULATION PROPORTION AND OF PROBABILITIES OF MISCLASSIFICATION FROM IMPERFECT BINOMIAL DATA BY COMPARISON WITH RESULTS OF A STANDARD METHOD OF KNOWN IMPERFECTION

Norman L. Johnson
University of North Carolina
Chapel Hill, North Carolina

and

Samuel Kotz
University of Maryland
College Park, Maryland

ABSTRACT: Observations from inspection by a 'test' method and a standard method are combined to provide estimators of population proportion, and of probabilities of misclassification for the test method. Results of Hochberg and Tenenbein (1983) and of Albers and Veldman (1984) are extended to the case where the standard method is not perfect, but its misclassification probabilities have known values. Both moment and maximum likelihood estimators are considered and some asymptotic properties of the resulting estimators are compared. Additional keywords: method of moments; binomial distribution; statistical differentials approximation (mathematics); bias; correlation; covariance.

ditto copy

Key Words and Phrases: Binomial distribution; information matrix; inspection errors; maximum likelihood; method of moments; statistical differentials.

1. INTRODUCTION

Suppose we have a large population, containing an unknown proportion, P, of individuals possessing a certain characteristic, which we will call 'nonconformance'. In a random sample, of size n, from this population, the distribution of the number, X, say, of nonconforming individuals will be binomial with parameters n,P so that

$$\Pr[X=x] = \binom{n}{x} P^x (1-P)^{n-x} \quad (x=0,1,\ldots,n).$$

We will represent this, symbolically, as

$$X \frown \mathrm{Bin}(n,P)\ , \quad \text{where } \frown \text{ denotes "is distributed as".}$$

If the individuals in the sample of size n are examined by an imperfect measuring device, which detects actual nonconformance with probability p, and (incorrectly) 'detects' nonconformance, when the individual is really not nonconforming, with probability p', then the distribution of Z, the number of individuals declared to be nonconforming, as a result of this inspection, will be binomial with parameters n, $Pp+(1-P)p'$. It is clear that the only parameter that can be estimated from observations on values of Z in independent samples is $Pp+(1-P)p'$.

Various methods have been suggested for obtaining data from which estimates of P, p and p' can be derived (e.g. Albers and Veldman (1984), Johnson and Kotz (1985)). Tenenbein (1970) suggested additional inspection of part of the sample by a perfect measuring device (for which p=1 and p'=0) and utilizing the resultant data. This method has been extended by Hochberg and Tenenbein (1983) to allow for inspection of a further sample, of size n_S, say, by the perfect measuring device (S).

In this paper, we study problems arising in this latter situation if the 'established' measuring device S is not perfect, but has known values p_S, p'_S for p,p' respectively. For convenience, we will denote the (unknown) values of p,p' for the measuring device under test (T) by p_T, p'_T respectively. We will also assume (when necessary) that $p_S \gg p'_S$ and $p_T \gg p'_T$.

2. ANALYSIS I (Moment Estimation)

As a consequence of the inspections we have the following sets of observations:

(i) n_S using S alone, with Z_S judged nonconforming (NC),

(ii) n_T using T alone, with Z_T judged NC,

(iii) n using both S and T, with results shown below:

S \ T	# NC	# not NC
# NC	Z_{11}	Z_{10}
# not NC	Z_{01}	Z_{00}

(# denotes 'number of'.) Evidently, $Z_{11}+Z_{10}+Z_{01}+Z_{00} = n$.

Under the assumption of random sampling from a population of effectively infinite size, we have that:

$$Z_S,\ Z_T \text{ and } \underset{\sim}{Z} = \begin{pmatrix} Z_{11} & Z_{10} \\ Z_{01} & Z_{00} \end{pmatrix} \text{ are mutually independent;} \tag{1.1}$$

$$Z_S \frown \mathrm{Bin}(n_S, \theta_S) \quad \text{with} \quad \theta_S = p_S P + p'_S(1-P)\ ; \tag{1.2}$$

$$Z_T \frown \mathrm{Bin}(n_T, \theta_T) \quad \text{with} \quad \theta_T = p_T P + p'_T(1-P)\ ; \tag{1.3}$$

$$\text{and} \qquad \underset{\sim}{Z} \frown \mathrm{Multinomial}\left(n\ ;\ \begin{pmatrix} \phi & \theta_S-\phi \\ \theta_T-\phi & 1-\theta_S-\theta_T+\phi \end{pmatrix} \right) \tag{1.4}$$

with $\phi = p_S p_T P + p_S' p_T'(1-P)$, where $\frown$ denotes "is distributed as".

Recall that p_S and p_S' have known values, and P is the (unknown) proportion of NC individuals in the population.

Hence

$$(n_S + n)\tilde{\theta}_S = Z_S + Z_{10} + Z_{11} \frown \mathrm{Bin}(n_S + n,\, \theta_S) \tag{2.1}$$

$$(n_T + n)\tilde{\theta}_T = Z_T + Z_{01} + Z_{11} \frown \mathrm{Bin}(n_T + n,\, \theta_T) \tag{2.2}$$

$$n\tilde{\phi} = Z_{11} \frown \mathrm{Bin}(n,\phi) \tag{2.3}$$

so that $\tilde{\theta}_S$ $\tilde{\theta}_T$ and $\tilde{\phi}$ (as defined in (2.1)-(2.3)) are unbiased estimators of θ_S, θ_T and ϕ respectively.

Hence $$\tilde{P} = (p_S - p_S')^{-1}(\tilde{\theta}_S - p_S') \tag{3.1}$$

is an unbiased estimator of P . Although the estimators

$$\tilde{p}_T = (\tilde{\theta}_S - p_S')^{-1}(\tilde{\phi} - p_S'\, \tilde{\theta}_T) \tag{3.2}$$

and $$\tilde{p}_T' = (p_S - \tilde{\theta}_S)^{-1}(p_S\tilde{\theta}_T - \tilde{\phi}) \tag{3.3}$$

are not unbiased estimators of p_T and p_T' respectively, the biases should not be large if sample sizes are adequate (see the example later in this section).

The variance-covariance matrix of the random variables in (2.1)-(2.3) is

$$\mathrm{Var}\left((n_S+n)\tilde{\theta}_S, (n_T+n)\tilde{\theta}_T, n\tilde{\phi}\right) = \begin{pmatrix} (n_S+n)\ {}_S(1-\theta_S) & n(\phi-\theta_S\theta_T) & n\phi(1-\theta_S) \\ n(\phi-\theta_S\theta_T) & (n_T+n)\theta_T(1-\theta_T) & n\phi(1-\theta_T) \\ n\phi(1-\theta_S) & n\phi(1-\theta_T) & n\phi(1-\phi) \end{pmatrix} \tag{4}$$

Hence (cf. (3.1))

$$\mathrm{Var}(\tilde{P}) = (n_S+n)^{-1}(p_S-p_S')^{-2}\theta_S(1-\theta_S) \tag{5.1}$$

and, using the method of statistical differentials (see, e.g. Johnson and Kotz (1969, Chapter 1, Section 7.5)) we obtain, after some algebraic manipulation:

$$\mathrm{var}(\tilde{p}_T) \doteq p_T^{\,2} P^{-2} (p_S - p_S')^{-2} \Big[\Big\{ n^{-1}\phi(1-\phi) - 2(n_T+n)^{-1} p_S' \phi (1-\theta_T) + (n_T+n)^{-1} p_S'^{\,2} \theta_T (1-\theta_T) \Big\} p_T^{-2}$$
$$-2(n_S+n)^{-1}\{\phi(1-\theta_S) - n(n_T+n)^{-1} p_S'(\phi - \theta_S\theta_T)\} p_T^{-1} + (n_S+n)^{-1}\theta_S(1-\theta_S) \Big] \quad (5.2)$$

An approximate expression for $\mathrm{var}(\tilde{p}_T')$ is obtained from (5.2) by replacing p_T by p_T' and P by (1-P), and interchanging p_S and p_S'.

An approximate formula for the bias of p_T is

$$E[\tilde{p}_T] - p_T \doteq p_T \left\{ \frac{\mathrm{var}(\tilde{\theta}_S)}{(\theta_S - p_S')^2} - \frac{\mathrm{cov}(\tilde{\theta}_S,\ \tilde{\phi} - p_S'\ \tilde{\theta}_T)}{(\theta_S - p_S')(\phi - p_S'\ \theta_T)} \right\} \quad (6)$$

which, after some reduction, gives a proportional bias (i.e. $100(\mathrm{bias})/p_T\%$)

$$\frac{100\{n_T p_S'(1-\theta_S) + n(1-p_S')\theta_S\}(\phi - \theta_S\theta_T)}{(n_S+n)(n_T+n)(\theta_S - p_S')^2(\phi - p_S'\theta_T)} \ \% \quad (7)$$

From (1.1)-(1.4)

$$\theta_S - p_S' = (p_S - p_S')P\ ; \quad \phi - p_S'\ \theta_T = p_T(p_S - p_S')P$$

and $\quad \phi - \theta_S\theta_T = P(1-P)(p_S - p_S')(p_T - p_T')$, so

the approximate proportional bias (7) is

$$\frac{100\{n_T p_S'(1-\theta_S) + n(1-p_S')\theta_S\}\ (1-P)(p_T - p_T')}{(n_S+n)(n_T+n)P^2(p_S - p_S')^2 p_T} \ \% \quad (7)'$$

which is positive and (since $p_T' < p_T$) less than

$$\frac{100\ G(1-P)}{(n_S+n)P^2(p_S - p_S')^2} \ \% \quad (8)$$

where

$$G = \frac{n_T}{n_T+n} p_S'(1-\theta_S) + \frac{n}{n_T+n}(1-p_S')\theta_S , \tag{9}$$

which lies between $p_S'(1-\theta_S)$ and $(1-p_S')\theta_S$.

Example. Using as 'typical' values of the probabilities p_S, p_S' and P the values 0.9, 0.1 and 0.1 respectively we find that

$$G = (n_T+n)^{-1}(0.082n_T + 0.162n)$$

(so that G lies between 0.082 and 0.162) and the approximate proportional bias of $\tilde{p}_T$ is between 0 and 1406.25 $G(n_S+n)^{-1}$% . Note that the upper limit is less than 227.8 $(n_S+n)^{-1}$%, so if $n_S + n > 100$ the approximate proportional bias is less than 2.28%. The next section contains a numerical assessment of formula (5.2), without specifying values of p_T and p_T'.

3. Some Numerical Approximations

Utilizing the reasonable assumption that $p_S \gg p_S'$, and neglecting terms in p_S' and $p_S'^2$ in the numerator of (5.2) we find

$$\mathrm{Var}(\tilde{p}_T) \doteqdot p_T^2 P^{-2}(p_S-p_S')^{-2}\left\{\frac{\phi(1-\phi)}{np_T^2} - \frac{2\phi(1-\theta_S)}{(n_S+n)p_T} + \frac{\theta_S(1-\theta_S)}{n_S+n}\right\} \tag{10}$$

Taking $p_S = 0.9$, $p_S' = 0.1$ so that $\theta_S = 0.8P + 0.1$ and $\phi = 0.9\, p_T P + 0.1\, p_T'(1-P)$, we obtain from (10)

$$\mathrm{Var}(\tilde{p}_T) \doteqdot \frac{p_T^2}{0.64P^2}\left[\frac{\{0.9p_TP + 0.1p_T'(1-P)\}\{1-0.9p_TP - 0.1p_T'(1-P)\}}{n\,p_T^2} + \frac{(0.8P+0.1)(0.9-0.8P) - \{1.8p_TP + 0.2p_T'(1-P)\}(0.9-0.8P)p_T^{-1}}{n_S + n}\right]$$

Now taking P = 0.1, we find

$$\operatorname{var}(\tilde{p}_T) \doteqdot \frac{p_T^{\,2}}{0.0064}\left[\frac{0.09(p_T+p_T')\{1-0.09(p_T+p_T')\}}{n\,p_T^{2}} - \frac{0.1476}{n_S+n}\cdot\frac{p_T'}{p_T}\right] \tag{11}$$

$$\doteqdot \frac{14.375}{n}\;(p_T+p_T')\{1-0.09(p_T+p_T')\} - \frac{23.06}{n_S+n}\cdot\frac{p_T'}{p_T} \tag{12}$$

Since $0.09(p_T+p_T')\{1-0.09(p_T+p_T')\} < \frac{1}{4}$ (because $0.09(p_T+p_T') < 1$) the right hand side of (6) is less than

$$(0.0256\,n)^{-1} < 39.1\;n^{-1}.$$

In the next section we will compare the asymptotic variances and covariances of $\tilde{\theta}_S, \tilde{\theta}_T$ and $\tilde{P}$ with those for maximum likelihood estimators $\hat{\theta}_S$, $\hat{\theta}_T$ and $\hat{P}$ of θ_S, θ_T and P respectively.

4. ANALYSIS II (Maximum Likelihood Estimators)

The likelihood function of Z_S, Z_T and Z is

$$\binom{n_S}{Z_S}\binom{n_T}{Z_T}\binom{n}{Z_{11},Z_{10},Z_{01},Z_{00}}\,\theta_S^{\,Z_S}(1-\theta_S)^{n_S-Z_S}\,\phi^{Z_{11}}(\theta_S-\phi)^{Z_{10}}(\theta_T-\phi)^{Z_{01}}(1-\theta_S-\theta_T+\phi)^{Z_{00}}$$

Equating derivatives of the log-likelihood to zero gives the following equations for $\hat{\theta}_S$, $\hat{\theta}_T$ and $\hat{\phi}$:

$$\frac{Z_S}{\hat{\theta}_S} - \frac{n_S-Z_S}{1-\hat{\theta}_S} + \frac{Z_{10}}{\hat{\theta}_S-\hat{\phi}} - \frac{Z_{00}}{1-\hat{\theta}_S-\hat{\theta}_T+\hat{\phi}} = 0 \tag{13.1}$$

$$\frac{Z_T}{\hat{\theta}_T} - \frac{n_T-Z_T}{1-\hat{\theta}_T} + \frac{Z_{01}}{\hat{\theta}_T-\hat{\phi}} - \frac{Z_{00}}{1-\hat{\theta}_S-\hat{\theta}_T+\hat{\phi}} = 0 \tag{13.2}$$

$$\frac{Z_{11}}{\hat{\phi}} - \frac{Z_{10}}{\hat{\theta}_S-\hat{\phi}} - \frac{Z_{01}}{\hat{\theta}_T-\hat{\phi}} + \frac{Z_{00}}{1-\hat{\theta}_S-\hat{\theta}_T+\hat{\phi}} = 0 \tag{13.3}$$

The information matrix is

$$\underset{\sim}{V}^{-1} = \begin{pmatrix} \dfrac{n_S}{\theta_S(1-\theta_S)} + \dfrac{n(1-\theta_T)}{(\theta_S-\phi)(1-\theta_S-\theta_T+\phi)} & \dfrac{n}{1-\theta_S-\theta_T+\phi} & -\dfrac{n(1-\theta_T)}{(\theta_S-\phi)(1-\theta_S-\theta_T+\phi)} \\ \dfrac{n}{1-\theta_S-\theta_T+\phi} & \dfrac{n_T}{\theta_T(1-\theta_T)} + \dfrac{n(1-\theta_S)}{(\theta_T-\phi)(1-\theta_S-\theta_T+\phi)} & -\dfrac{n(1-\theta_S)}{(\theta_T-\phi)(1-\theta_S-\theta_T+\phi)} \\ -\dfrac{n(1-\theta_T)}{(\theta_S-\phi)(1-\theta_S-\theta_T+\phi)} & -\dfrac{n(1-\theta_S)}{(\theta_T-\phi)(1-\theta_S-\theta_T+\phi)} & \dfrac{n\{\theta_S\theta_T(1-\theta_S-\theta_T)+2\theta_S\theta_T\phi-\phi^2\}}{\phi(\theta_S-\phi)(\theta_T-\phi)(1-\theta_S-\theta_T+\phi)} \end{pmatrix} . \tag{14}$$

The determinant is

$$|\underset{\sim}{V}^{-1}| = \frac{n}{\phi(\theta_S-\phi)(\theta_T-\phi)(1-\theta_S-\theta_T+\phi)} \left[\frac{n_S n_T \gamma}{\theta_S\theta_T(1-\theta_S)(1-\theta_T)} + n(n_S+n_T+n) \right]$$

with $\gamma = \theta_S\theta_T(1-\theta_S-\theta_T+\phi) - \phi(\phi-\theta_S\theta_T)$. (15)

and from the asymptotic variance-covariance matrix $\underset{\sim}{V}$ we obtain

$$\mathrm{Var}(\hat{\theta}_S) \doteq \frac{n}{|\underset{\sim}{V}^{-1}|} \left\{ \frac{n_T}{\theta_T(1-\theta_T)} \left(\frac{1}{\phi} + \frac{1}{\theta_S-\phi} + \frac{1}{\theta_T-\phi} + \frac{1}{1-\theta_S-\theta_T+\phi} \right) + n \left(\frac{1}{\theta_T-\phi} + \frac{1}{1-\theta_S-\theta_T+\phi} \right) \left(\frac{1}{\phi} + \frac{1}{\theta_S-\phi} \right) \right\}$$

$$\doteq \theta_S(1-\theta_S)\ (\gamma n_S^{-1} + \delta N^{-1})/(\gamma + \delta)$$

where $N = n_S+n_T+n$ (= total number of observations) and

$$\delta = \frac{nN}{n_S n_T} \theta_S\theta_T(1-\theta_S)(1-\theta_T)$$

The MLE of P is

$$\hat{P} = (p_S-p_S')^{-1}(\hat{\theta}_S-p_S') \tag{16}$$

The asymptotic efficiency of $\tilde{P}$ (see (3.1)) is the same as that of $\tilde{\theta}_S$, which is

$$100(n_S+n)(\gamma\ n_S^{-1} + \delta N^{-1})/(\gamma+\delta)\ \% \tag{17}$$

Taking $p_S=p_T=0.9$, $p'_S=p'_T=0.1=P$ and $n_S=n_T=n$ $(=\frac{1}{3}N)$ we find $\gamma = 0.0184680$ and $\delta = 0.0653573$, so (17) becomes

$$100(n_S+n)\ (0.2203\ n_S^{-1} + 0.7797\ N^{-1})$$

$$= 2(0.2203 + 0.2599) = 96.04\% .$$

The asymptotic variance of the MLE $\hat{\phi}$ is

$$\mathrm{var}(\hat{\phi}) \doteqdot \frac{1}{n}\,\frac{\delta\phi(1-\phi) - \delta\phi^2N^{-1}\{n_S\theta_S^{-1}(1-\theta_S) + n_T\theta_T^{-1}(1-\theta_T)\} + \phi(\theta_S-\phi)(\theta_T-\phi)(1-\theta_S-\theta_T+\phi)}{\delta+\gamma} \tag{18}$$

On the other hand, recalling that $\mathrm{var}(\tilde{\phi}) = n^{-1}\phi(1-\phi)$, we find for the numerical values of the parameters used above, that the asymptotic efficiency of the moment estimator ϕ is

$$100 \times \frac{0.0653573 \times 0.09\{0.91 - (2/3) \times 0.09 \times (0.18)^{-1} \times 0.82\} + 0.09 \times 0.09^2 \times 0.73}{0.09 \times 0.91\ 0.0653573 + 0.0184680)}$$

$$= 100 \times \frac{0.0037449 + 0.0005322}{0.0068653} = 62.30\%$$

The markedly lower asymptotic efficiency of $\tilde{\phi}$ is associated with the fact that it does not utilize the information on values of θ_S and θ_T which is available from the other (n_S+n_T) observations. Some support for this statement comes from the asymptotic efficiency of $\tilde{\phi}$ if the values of θ_S and θ_T are <u>known</u>. This is

$$100 \times \frac{(\theta_S-\phi)(\theta_T-\phi)(1-\theta_S-\theta_T+\phi)}{\gamma(1-\phi)} \tag{19}$$

With the numerical values of θ_S, θ_T and ϕ which we have been using above this would give an asymptotic efficiency of only <u>35.18</u>%.

5. CORRELATIONS

The asymptotic covariances between the MLE's $\hat{\theta}_S$ and $\hat{\theta}_T$, and between $\hat{\theta}_S$ and $\hat{\phi}$, are

$$\mathrm{cov}(\hat{\theta}_S,\hat{\theta}_T) \doteq N^{-1}\,\delta(\phi-\theta_S\theta_T)/(\gamma+\delta) \tag{20}$$

and
$$\mathrm{cov}(\hat{\theta}_S,\hat{\phi}) \doteq N^{-1}\,\delta\,\phi\{nm_T^{-1}(1-\phi\theta_S^{-1}) + 1-\theta_S\}/(\gamma+\delta) \tag{21}$$

Since $0 \le \theta_S\theta_T \le \phi \le \theta_S \le 1$, the covariances cannot be negative. It is of interest to compare (20) with the covariance between the moment estimators $\tilde{\theta}_S$ and $\tilde{\theta}_T$, which is

$$\mathrm{cov}(\tilde{\theta}_S,\tilde{\theta}_T) = n(n_S+n)^{-1}(n_T+n^{-1}(\phi-\theta_S\theta_T) \tag{22}$$

and (21) with the covariance between $\tilde{\theta}_S$ and $\tilde{\phi}$, which is

$$\mathrm{cov}(\tilde{\theta}_S,\tilde{\phi}) = (n_S+n)^{-1}\phi(1-\theta_S) \tag{23}$$

Using our illustrative parameter values ($p_S=p_T=0.9$; $p'_S=p'_T=0.1=P$; $n=n_S=n_T$) we obtain the following values for correlations:

$$\mathrm{corr}(\tilde{\theta}_S,\tilde{\theta}_T) = 0.195; \quad \mathrm{corr}(\tilde{\theta}_S,\tilde{\phi}) = 0.475$$
$$\mathrm{corr}(\hat{\theta}_S,\hat{\theta}_T) \doteq 0.021; \quad \mathrm{corr}(\hat{\theta}_S,\hat{\phi}) \doteq 0.513$$

Note that if $n=0$ (and n_S, $n_T > 0$) then

$$\mathrm{corr}(\tilde{\theta}_S,\tilde{\theta}_T) = \mathrm{corr}(\hat{\theta}_S,\hat{\theta}_T) = 0\,.$$

Also, if $n_S=n_T=0$ (and $n>0$) then

$$\mathrm{corr}(\tilde{\theta}_S,\tilde{\theta}_T) = \mathrm{corr}(\hat{\theta}_S,\hat{\theta}_T) = (\phi-\theta_S\theta_T)\{\theta_S\theta_T(1-\theta_S)(1-\theta_T)\}^{-\frac{1}{2}}$$
$$(= 0.390)$$

and

$$\mathrm{corr}(\tilde{\theta}_S,\tilde{\phi}) = \mathrm{corr}(\hat{\theta}_S,\hat{\phi}) = [\phi(1-\theta_S)/\{\theta_S(1-\phi)\}]^{\frac{1}{2}}$$
$$(= 0.147)$$

(Numbers in parentheses are values corresponding to $p_S=p_T=0.09$, $p'_S=p'_T=0.1=P$.

Acknowledgement

Dr. Samuel Kotz's work was supported by the U.S. Office of Naval Research under contract N00014-84-K-0301.

REFERENCES

Albers, W. and Veldman, H.J. (1984) Adaptive estimation of binomial probabilities under misclassification, Statist. Neerland., 38, 233-247.

Hochberg, Y. and Tenenbein, A. (1983) On triple sampling schemes for estimating from binomial data with misclassification errors, Commun. Statist. - Theor. Math., 12, 1523-1533.

Johnson, N.L. and Kotz, S. (1969) Distributions in Statistics: Discrete Distributions, Wiley, New York.

Johnson, N.L. and Kotz, S. (1985) Some tests for detection of faulty inspection, Statist. Hefte, 26, 19-29.

Tenenbein, A. (1970) A double sampling scheme for estimating from binomial data with misclassification, J. Amer. Statist. Assoc., 65, 1350-1362.

NONCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AD-A159 926

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Nonclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) UM/MSS/85/7	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Univ. of Maryland 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION U.S. Office of Naval Research
6c. ADDRESS (City, State and ZIP Code) Dept of Management and Statistics University of Maryland College Park, MD 20742	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION ONR 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N 00014-84-K-301
8c. ADDRESS (City, State and ZIP Code) Stat. and Probability Program Office of Naval Research Arlington, VA 22217	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO.

11. TITLE (Include Security Classification) "Estimation of population proportion and of probabilities of misclassification from imperfect binomial data by comparison with results of a standard method of known imperfection"

12. PERSONAL AUTHOR(S) JOHNSON, N.L. and KOTZ, Samuel

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
TECHNICAL	FROM 9/84 TO 9/85	August 1985	11 pages

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	binomial distribution; information matrix; inspection errors; maximum likelihood; method of moments; statistical differentials.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Observations from inspection by a 'test' method and a standard method are combined to provide estimators of population proportion, and of probabilities of misclassificatinn for the test method. Results of Hochberg and Tenenbein (1983) and of Albers and Veldman (1984) are extended to the case where the standard method is not perfect, but its misclassification probabilities have known values. Both moment and maximum likelihood estimators are considered and some asymptotic properties of the resulting estimators are compared.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION Nonclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Samuel Kotz	22b. TELEPHONE NUMBER (Include Area Code) (301) 454-6805 22c. OFFICE SYMBOL

DD FORM 1473, 83 APR   EDITION OF 1 JAN 73 IS OBSOLETE.

NONCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

END

FILMED

11-85

DTIC